Answer Referee #1

General comments to both referees:

Most of the suggestions received have been adopted. From this, major changes have been made on the study, comprising: (1) the inclusion of Gini importance plots; (2) a modification of the set of predictive variables, adding the boundary layer height and the total cloud cover and (3) the normalization of the meteorological variables. On this basis, we build a new model to decouple the effect of the meteorology for the analysis of the relative changes during COVID-19 period. Although these suggestions led to an overall better implementation of the Random Forest model and a better subsequent analysis, the conclusions remain almost unchanged. Nevertheless, with the new estimated values for the relative changes, we will include substantive modifications in section 3.1.

<table>
<tr><td colspan="2">This study was majorly inspired by Grange et al 2019 to use a random forest model (RF) to train, validate and predict the air quality concentration in a megacity of Argentina. Although the methodology is not new, this study has the potential to have a significant improvement to fill the data gap given the reason that some monitored trace gas concentrations were lacking but the pollutants are becoming a concern for local authorities. (line 145- 150) I would like to propose a few major revisions for the authors of this manuscript. After done with that improvement, I think it would be strong enough to publish on earth system science data:</td></tr>
<tr><td>1. Random forest models indeed are easy and quick to train. But if this study is only focusing on predicting the time series of pollution during COVID-19, there are better and more efficient machine learning models such as ARIMA compared with random forest. The key reason for many studies that authors mentioned in this article used RF is because it could provide key components reflecting the non-linear relationship among emissions, chemical reaction, and meteorological effects. Please see figure 3 in Grange et al 2019 and figure 1 in Yang et al. It is easy to generate that Gini importance either through python or R. Since this code the authors used were based on R, I would suggest they follow the</td><td>Following your suggestion, we generated the Gini importance analysis using the ranger package in R for all the variables considered in the model. As you mentioned, these results were helpful for us to understand the underlying role (non-linear effects) of some variables in the concentrations of the pollutants analyzed. As an example, Gini plots reveal the importance to consider the boundary layer height as an explanatory variable, particularly for CO and NO description. On the basis of this analysis, we modified the set of the explanatory variables chosen for each model, adding the boundary layer height (blh). For example, for the CNEA site: (1) for CO the set of explanatory variables was change from {t2; rh2; U; V; gasoline diurnal cycle} to { t2; rh2; sea level pressure; ws; wd; blh; total cloud cover; gasoline diurnal cycle; hour} and (2) for NO the set was modified from {t2; rh2; U; V; gasoline and diesel diurnal cycles} to {t2; rh2; sea level pressure; ws; wd; blh; total cloud cover; gasoline and diesel diurnal cycle; CO;hour; month, weekday}

Regarding the aim of the study, note that we presented 3 different goals. Two of them (lines 65-68) are: (1) the prediction of the time series during COVID-19 and (2) the development of a model for air quality forecasts for the Metropolitan Area of Buenos Aires at a low computational cost. In line 74 and 75, we also explain that we are</td></tr>
</table>

	code from Grange et al 2019 to generate the Gini importance plot. After getting those plots, you can compare your RF results with the reasons from previous literature to see if it makes sense.	providing (3) the first O_3 and SO_2 observational datasets in more than a decade in Buenos Aires. However, from your comment, we realized that (2) and (3) were not highlighted in the abstract. Therefore, we will include some modifications in the abstract.
2.	The authors also mentioned several times the meteorological impacts on air pollution. The earliest signal has been seen during COVID was the study in China from Le et al, 2020 which authors may consider mentioning this study result. Then it would also be beneficial to consider using RF models to generate new predictions by normalizing meteorological factors. This would give the third line in each panel of figure 3. You can do this by following the methodology in figure 5 of Grange et al 2019 and figure 2 of Yang et al 2021. Vu et al 2019 made an additional improvement to weather normalization which you may also consider using this methodology.	We decided to adopt this suggestion making the following changes to our study: (1) Building a new RF model, normalizing the weather variables, adopting the Shi et al 2021 approach (which follows Vu et al 2019 not normalizing time variables, and Grange et al 2019 resampling from the whole study period). With this new model, we re-estimate the relative changes between MAM 2019 and MAM 2020, but using the normalized concentrations. This new approach allowed us to re-analyze the effects of the emissions changes during COVID-19 period, decoupling the effects of meteorological conditions. These new results will be included in the revised version of the manuscript. (2) Leaving the previous RF model as a predictive tool for air quality forecast in Buenos Aires, but adding new explanatory variables, as explained in Answer #1. This approach allowed us to assess the combined effect of the particular meteorological situation during COVID-19 restriction period, and the reduction in the emissions that actually occurred. The comparison between observations and concentrations that would have occurred under normal emissions conditions (BAU scenario) estimated with this RF model will be kept in the revised version of the manuscript. To explain these methodological changes Figure 2 in the revised version will be replaced by Figure 2 (revised) presented further down. In addition, we will expand our references, including Le et al, 2020, Vu et al 2019, Shi et al 2021 and Grange et al 2019

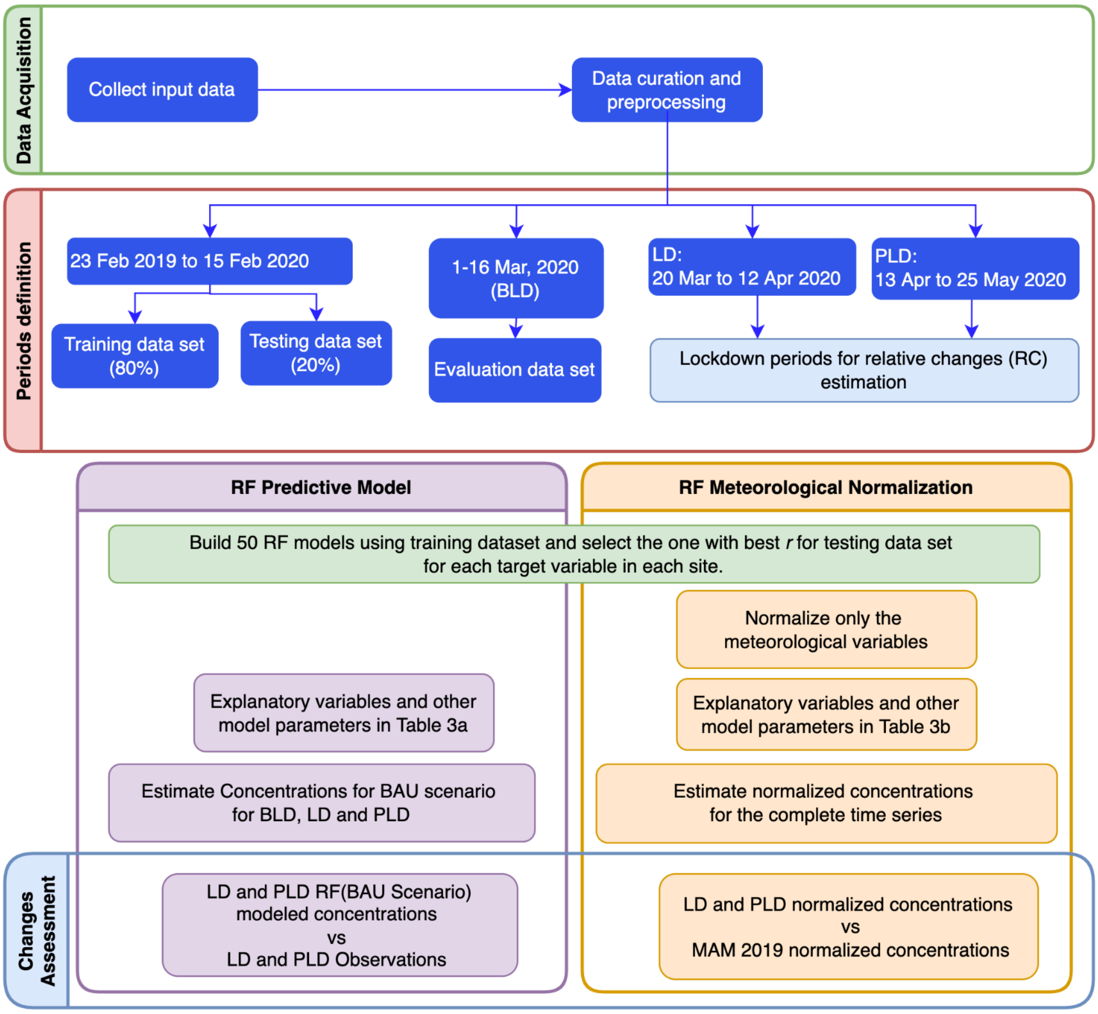

Figure 2 (revised)

3. The description of details of why and how to interpret bivariate polar plots here is vague. The reason that Grange 2019 used this bivariate polar plot is that they wanted to prove that wind influences the dispersion of pollutants. Therefore it's better to show the meteorological impacts from the above suggestions that I mentioned first. Then applying a bivariate polar plot by combining with winds components if the meteorological impact is the dominant factor here.	Even though we knew that wind speed and direction were key variables for pollutant dispersion in Buenos Aires, we used bivariate plots not for proving that, but as a tool for source location, detection and characterisation (following Grange et al 2016 and Carslaw 2013). We think that these graphs provide a graphical support for showing joint wind speed, wind direction dependence of air pollutant concentrations under different scenarios, in this case with and without COVID-19 restrictions. We used these plots to gain knowledge on potential sources that were missing during COVID-19 lockdown periods, such as the impact on the different pollutants of reducing vehicular emissions. In order to improve the way to interpret these useful plots, we will modify the line 228, adding the following sentence at the end: "These should help us detect, locate and characterize pollution sources, as done by Carslaw *et al.* 2013 and by

	Grange *et al.* 2016." *Stuart K. Grange, Alastair C. Lewis, David C. Carslaw, Source apportionment advances using polar plots of bivariate correlation and regression statistics, Atmospheric Environment, Volume 145, 2016.* *Carslaw 2013, David C. Carslaw, Sean D. Beevers, Characterizing and understanding emission sources using bivariate polar plots and k-means clustering*
4. For the relationship between NO_2 and diesel please refer to Yang et al 2021. You can also consider normalizing other anthropogenic factors besides meteorology based on the result from the first suggestion.	Thanks for your suggestion. Unfortunately we couldn't consider other anthropogenic factors because there is no available traffic data in Buenos Aires, with the temporal and geographical disaggregation needed.
5. Please indicate how much data by # not percentage is used for training and how much data is used for validation/prediction. Due to some restrictions of the monitoring campaign, please indicate how the authors dealt with inadequate data for specific variables.	The CNEA site has 9198 valid data points but only 7150 of those were before the BLD period (please, refer to Figure 2 revised for period definitions); 80% of these (5720) were used to train the model and the rest (1430) for testing. The independent period for evaluation (namely BLD) is composed of 360 data points. There are 4 days between BLP and LD that were not taken into account, because they have been considered as a "transition period". The numbers for the Parque Centenario site are: 8710 data points before the BLD period, with 6968 used for training and the rest for testing (1742). Inadequate data was considered as missing data, and was not replaced. The revised version of the manuscript will clarify this issue, modifying the lines 188-192.
6. The clarity and context need significant improvement to better draw out why the results are significant.	We tried to emphasize in the conclusions the importance of our results, including: A. The importance of producing novel SO_2 and O_3 data, in a basin with lack of monitoring data for these pollutants in residential/commercial areas (the only data available corresponds to an industrial area). It is well-known the importance of having non-industrial air quality data for air quality model validation. B. The importance of having a tool for air quality forecasts in Buenos Aires at a low computational cost, which could be useful for air quality management in the city. C. The analysis of the effects of COVID restrictions in air quality, analyzing (as was made in the rest of the world) the effects in

	primary pollutants reduction, but also in O_3 increase. D. The probable VOCs limited regime for the Buenos Aires Atmosphere. As was said in our Answer#1, the revised version of our work will include a change in the abstract reflecting all these issues , but we will also include modifications in the conclusions to better highlight the relevance of our results.
The following paper should have been referenced and discussed in the manuscript: Grange, Stuart K., and David C. Carslaw. "Using meteorological normalisation to detect interventions in air quality time series." Science of the Total Environment 653 (2019): 578-588. Yang, Jiani, Yifan Wen, Yuan Wang, Shaojun Zhang, Joseph P. Pinto, Elyse A. Pennington, Zhou Wang et al. "From COVID-19 to future electrification: Assessing traffic impacts on air quality by a machine-learning model." Proceedings of the National Academy of Sciences 118, no. 26 (2021). Vu, Tuan V., Zongbo Shi, Jing Cheng, Qiang Zhang, Kebin He, Shuxiao Wang, and Roy M. Harrison. "Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique." Atmospheric Chemistry and Physics 19, no. 17 (2019): 11303-11314. Le, Tianhao, Yuan Wang, Lang Liu, Jiani Yang, Yuk L. Yung, Guohui Li, and John H. Seinfeld. "Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China." Science 369, no. 6504 (2020): 702-706.	Thank you for this suggestion, we will add these references to the manuscript.